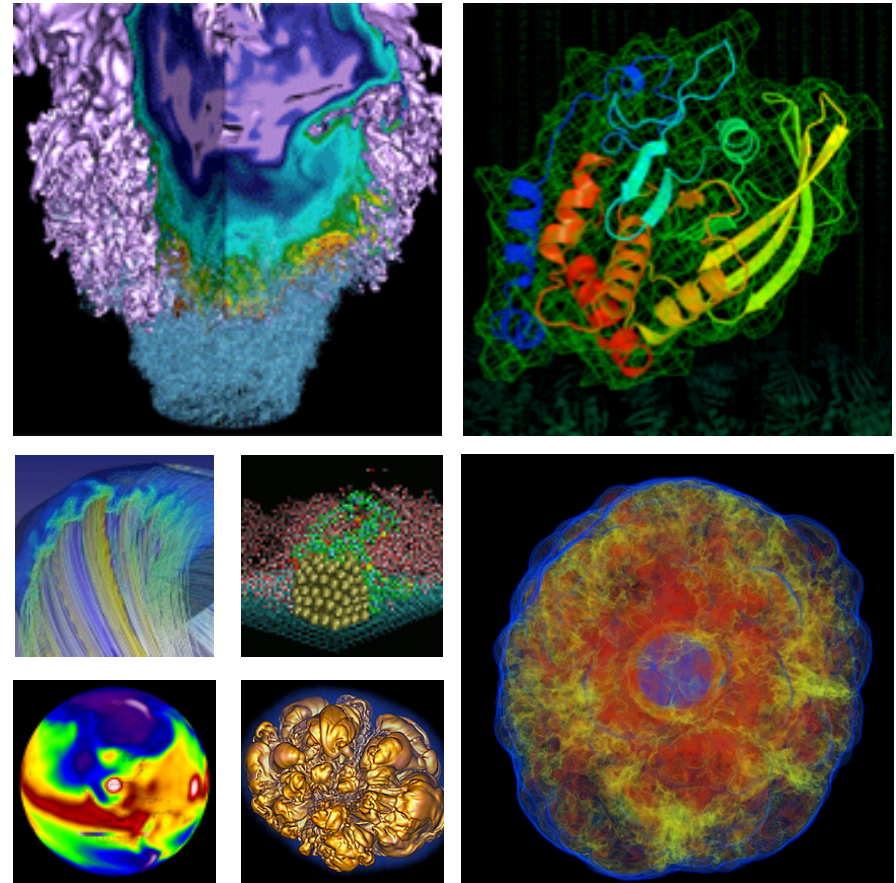


# Transferring Data at NERSC



**Jeff Porter & Shreyas Cholia**  
**Data and Analytics Services**  
**New User Training 2017**

February 24<sup>th</sup>, 2017

# Dedicated Data Transfer System: Data Transfer Nodes



- **Data Transfer Nodes (DTN) are dedicated servers for moving data at NERSC.**
  - Servers include high-bandwidth network interfaces & are tuned for efficient data transfers
    - Monitored bandwidth capacity between NERSC & other major facilities such as ORNL, ANL, BNL, SLAC...
  - Provide direct access to global NERSC file systems
  - Can be used to move data between NERSC systems & HPSS
- **Use the NERSC DTNs to move large volumes of data in and out of NERSC or between NERSC systems**

- **The recommended tool for moving data in & out of NERSC**
  - <http://www.globus.org/>
  - Reliable & easy-to-use web-based service:
    - Automatic retries
    - Email notification of success or failure
  - Accessible to all NERSC users
  - NERSC managed endpoints for optimized data transfers
    - Configured for general & specific uses
- **Globus extensive documentation covers**
  - Web based interaction with service
  - REST/API for scripted interactions with service
  - Globus Connect Server & Personal for setting up additional remote endpoints such as your personal laptop

# NERSC Managed Globus Endpoints



- **Available for all NERSC users**
  - Maps data servers to NERSC resources
  - See: <http://www.nersc.gov/users/storage-and-file-systems/transferring-data/globus-online/>

## NERSC Endpoints

For most use cases →

For special uses {

Now obsolete {

| Endpoint Name | Description  | Recommended Use  |
|---------------|--|--|
| NERSC DTN     | Multi-node, high performant transfer system with access to all NERSC Global File systems (NGF) as well as the large Cori Scratch | Almost all data transfers needs into & out of NERSC        |
| NERSC HPSS    | Single node system connected directly to the NERSC HPSS tape archive   | Remote transfers into & out of HPSS                        |
| NERSC Edison  | Single node system connected to NGF and uniquely to the Edison scratch file system   | <i>Only recommended for access to Edison scratch</i>       |
| NERSC PDSF    | Single node system connected to NGF and the two remaining PDSF-specific file systems, eliza3 and eliza18                         | <i>Only recommended for access to /eliza3 and /eliza18</i> |
| NERSC Cori    | Originally a dual-node system needed for accessing the Cori scratch file system. The endpoint is the same as NERSC DTN           | <i>Use NERSC DTN instead</i>                               |
| NERSC DTN-JGI | Single node system that was used to access JGI-specific file systems, which are now connected to the NERSC DTN servers.          | <i>Use NERSC DTN instead</i>                               |

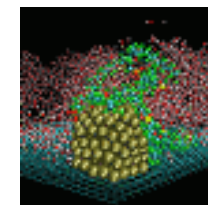
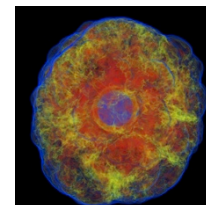
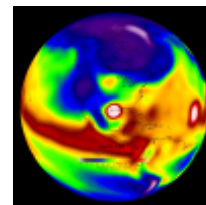
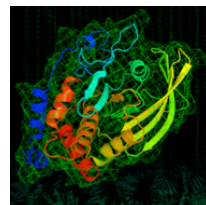
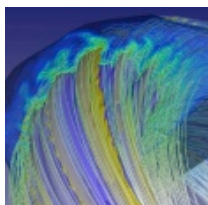
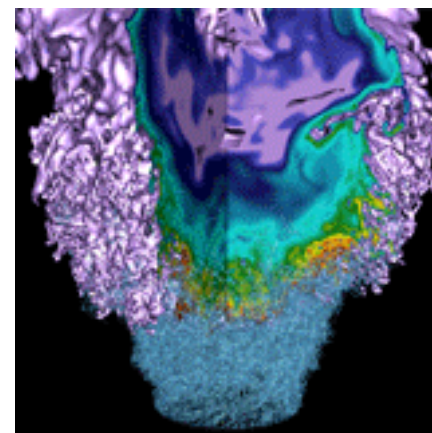


U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Globus: Quick Introduction



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Logging into Globus



Log in to use Globus Web App

Use your existing organizational login

e.g. university, national lab, facility, project, Google or [Globus ID](#)

(Your Globus username and password used prior to February 13, 2016 is now Globus ID)

NERSC

Continue

You can log into Globus Online with your NERSC Username & Password

[oauth.nerisc.gov](https://oauth.nerisc.gov)

Welcome to the OAuth for MyProxy Client Authorization Page. The Client below is requesting access to your account. If you approve, please sign in with your username and password.

## Client Information

Name: Globus  
URL: <https://www.globus.org>

Username

Password

Sign In

Cancel



# Endpoint Selection



- Search in the Globus Online Web application:

**Manage Endpoints**

Endpoint List

+ add Globus Connect Personal endpoint + add Globus Connect Server endpoint

recently used in use shared with me shared by me administered by me search all

NERSC

| endpoint  | status     | credential         |  |  |
|---|------------|--------------------|--|--|
| <b>NERSC Cori</b><br>Managed Public Endpoint    | not active | activate           |  |  |
| <b>NERSC DTN</b><br>Managed Public Endpoint     | ready      | expires in 12 days |  |  |
| <b>NERSC DTN JGI</b><br>Managed Public Endpoint | not active | activate           |  |  |
| <b>NERSC Edison</b><br>Managed Public Endpoint  | not active | activate           |  |  |
| <b>NERSC HPSS</b><br>Managed Public Endpoint    | not active | activate           |  |  |
| <b>NERSC PDSF</b><br>Managed Public Endpoint    | not active | activate           |  |  |

# Activating a NERSC endpoint



## Transfer Files

Endpoint

Path

Please authenticate to access this endpoint

Login Server

Username

Password

▼ advanced

Activate endpoint with your  
NERSC Username & Password

Advanced option:  
select up to 277 hours  
lifetime for activation

## Transfer Files

Endpoint

Path

Please authenticate to access this endpoint

Login Server

Username

Password

▲ advanced

Credential Lifetime

## Transfer Files

Endpoint

Path

|           |        |
|-----------|--------|
| NERSC     | Folder |
| PDSF      | Folder |
| STAR      | Folder |
| alicework | Folder |
| bin       | Folder |
| bosco     | Folder |
| dataTran  | Folder |
| davinci   | Folder |

Active endpoint opens  
file system browser



# Transfer Data From NERSC to another endpoint ... even your laptop



## Transfer Files

Endpoint: NERSC DTN Path: /~/NERSC/data/ Go

Endpoint: porter#laptop Path: /~/Documents/NERSC/Data Representat Go

select none up one folder refresh list

|                      |           |
|----------------------|-----------|
| capacity_factors.hdf | 141.41 MB |
|----------------------|-----------|

select all up one folder refresh list

|   |           |
|---|-----------|
| 2015_03_Qserv_NERSC.pdf                       | 1.83 MB   |
| DBs-NERSC.pptx                                | 1.21 MB   |
| DataStorageandRepresentationWorkingGroup.docx | 37.84 KB  |
| MongoDB 3.0 - LBL.pdf                         | 1.27 MB   |
| MongoDB 3.0 - LBL.pptx                        | 924.01 KB |
| esnet-cassandra-2015.pdf                      | 10.58 MB  |
| esnet-cassandra-2015.pptx                     | 10.13 MB  |

Laptop endpoint:  
Globus connect personal

select file

Transfer initiated

## Transfer Files

Transfer request submitted successfully. Task id: d9966c48-f999-11e6-ba9c-22000b9a448b

Endpoint: NERSC DTN Path: /~/NERSC/data/ Go

Endpoint: porter#laptop Path: /~/Documents/NERSC/Data Representat Go

select none up one folder refresh list

select all up one folder refresh list

|        |         |
|--------|---------|
| SC.pdf | 1.83 MB |
|        | 1.21 MB |

Activity

Task List

Transfer complete

✓ NERSC DTN to porter#laptop  
transfer completed 21 minutes ago

Overview

Event Log

Task ID d9966c48-f999-11e6-ba9c-22000b9a448b

Files 1

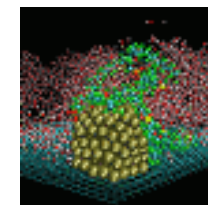
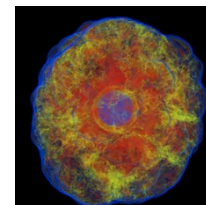
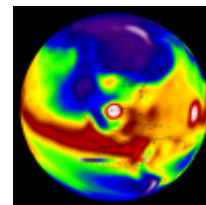
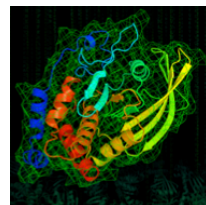
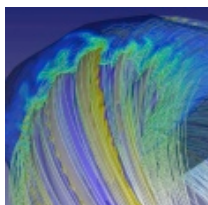
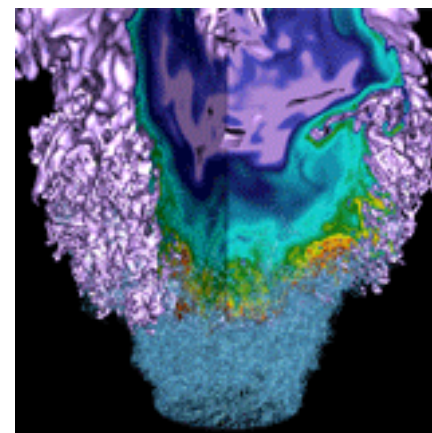


U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# Other tools and considerations



U.S. DEPARTMENT OF  
**ENERGY**

Office of  
Science



# NERSC Data Transfer Nodes for direct access



- **NERSC maintains several DTNs with login access**
  - dtn01.nersc.gov, dtn02, dtn03, dtn04
- **Familiar module environment**
  - Limited software deployment for data transfer needs
- **Global file systems are available**
  - /global/project, /global/projecta, /global/cscratch1
  - Excludes some Edison & PDSF specific systems
    - Edison: /scratch1, /scratch2, /scratch3
    - PDSF: /eliza3, /eliza18

# Transferring with NERSC HPSS



- **HPSS tape archive is recommended for storing/archiving large amounts of data and/or for long periods of time**
  - See: <http://tinyurl.com/nerschpss>
- **Use interactive DTNs with hsi/htar to move data to/from HPSS and NERSC file systems**
  - HSI for individual files and conditional access
  - HTAR for aggregation & optimization of storage/archival
- **For archival & retrieval of large number of files**
  - Aggregate data into smaller number of files with 'htar' for storing
  - Retrieve many files using tape-ordered routines:
    - <http://www.nersc.gov/users/storage-and-file-systems/hpss/storing-and-retrieving-data/mistakes-to-avoid/>
- **Also use Globus Online: NERSC HPSS endpoint**
  - However Globus does not directly support aggregation with 'htar' or tape-ordering
  - Preferred use is for small number of large files

# General Tips

---



- **Use Globus Online for large, automated or monitored transfers**
- **scp is fine for smaller, one-time transfers (<100MB)**
  - But note that Globus is also fine for small transfers
- **Don't use DTN nodes for non-data transfer purposes**
  - ever system has login nodes for more general routine tasks
- **Plain “cp” is still used for transfers within file systems**

# Performance Considerations



- **Performance is often limited by the remote endpoint**
  - Not tuned for WAN transfers or have limited network link
  - These can lower performance < 100 MB/sec.
- **File system contention may be an issue**
  - Try the transfer at a different time or on a different FS.
- **Don't use your \$HOME directory**
  - Instead use \$SCRATCH, /global/project, ...
- **If you think you are not getting the performance you expect, let us know: [consult@nersc.gov](mailto:consult@nersc.gov)**



# Access for External Collaborators

---



- **NERSC supports project-level public http access**
  - Project specific area can be created:
    - /global/project/projectdirs/<yourproject>/www
    - Owned by project PI or designate
  - Available for public access
    - <http://portal.nersc.gov/project/<yourproject>>

# For more information

---



- **Data transfer info**
  - <http://www.nersc.gov/users/data-and-file-systems/transferring-data>
- **General DTN info**
  - <http://www.nersc.gov/systems/data-transfer-nodes/>
- **Feedback / Problems?**
  - [consult@nersc.gov](mailto:consult@nersc.gov)
- **Globus Support**
  - <https://www.globus.org/support/>